Advancing Personality Type Prediction: Utilizing Enhanced Machine and Deep Learning Models with the Myers-Briggs Type Indicator

Mohammad Hossein Amirhosseini
Department of Computer Science and Digital Technologies
University of East London
London, United Kingdom
m.h.amirhosseini@uel.ac.uk

Amin Karami
Department of Computer Science and Digital Technologies
University of East London
London, United Kingdom
a.karami@uel.ac.uk

Fatima Kalabi
Homerton University Hospital
London, United Kingdom
f.kalabi@nhs.net

***Abstract*—This study addresses key gaps in personality prediction research by conducting a comprehensive comparison of machine learning and deep learning models on a new, large dataset of MBTI personality types. Previous studies predominantly focused on the Big Five framework and overlooked MBTI due to limited datasets. Moreover, basic hyperparameter tuning techniques, label imbalance, and insufficient text lengths in training samples have constrained the accuracy and generalizability of past models. To address these issues, this research employs a large balanced MBTI dataset with sufficient text lengths and optimizes models using Bayesian optimization. Models compared include Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVC), Naïve Bayes, LightGBM, XGBoost, Multilayer Perceptron (MLP), and Bidirectional Encoder Representations from Transformers (BERT). Results demonstrate that deep learning models outperform traditional methods, with BERT achieving the highest accuracy (93%), followed by XGBoost (86%) and SVC (85%). The BERT model also significantly outperformed the models implemented in previous works in this field. This work provides actionable insights into model selection and optimization, showcasing the utility of advanced techniques like Bayesian optimization in enhancing predictive performance. By addressing these gaps, the study lays the foundation for robust, scalable personality prediction models applicable in psychology, career counselling, and personalized marketing.**

Keywords—Deep Learning, Machine Learning, Large Language Model, Bayesian Optimization, Personality Prediction

I. Introduction

Understanding personality has historically been a cornerstone of psychological research, providing critical insights into human behavior and interpersonal interactions [1]. Among the tools for assessing personality, the Myers-Briggs Type Indicator (MBTI) stands as one of the most recognized frameworks, rooted in Jungian theory. The framework categorises people into sixteen separate personality profiles, each characterised by specific behavioural tendencies, cognitive styles, and emotional responses [2]. However, while the MBTI has found widespread application in different fields, traditional assessment methods face challenges related to scalability, efficiency, and the dynamic applicability of their insights [3].

In recent times, the integration of machine learning has transformed the landscape of personality psychology, offering significant improvements in the precision, scalability, and contextual adaptability of personality evaluations. By harnessing expansive datasets and leveraging complex algorithms that can interpret a wide array of behavioural inputs, machine learning supports a data-centric framework for predicting personality traits. Although this technological evolution presents certain complexities, it also opens avenues for personalized innovations across various domains, including adaptive learning environments, precision-targeted marketing initiatives, and the strategic formation of high-performing teams in organisational contexts.

This study's literature review explores the diverse applications of machine learning in the domain of personality classification, critically analysing its effectiveness while pinpointing existing research gaps and future possibilities for progression in this area. Through an integrated review of both foundational research and cutting-edge methodologies, this section presents a holistic understanding of the current advancements and charts potential directions for the evolution of personality prediction using machine learning technologies.

Building on these insights and after thoroughly assessing the gaps in current research, the study employed a newly compiled, extensive dataset and formulated a comprehensive machine learning methodology to predict personality types as defined by the Myers-Briggs Type Indicator (MBTI). Model performance was enhanced through Bayesian optimization techniques, ensuring finely tuned hyperparameters. Subsequently, models were evaluated using a suite of metrics, including accuracy, F1 score, precision, and recall. A comparative evaluation was also performed to measure model effectiveness, ultimately identifying the most accurate and reliable model for personality classification.

II. Previous Works on Personality Prediction

A. Machine Learning and Deep Learning Approaches

Recent research has employed diverse natural language processing (NLP) and machine learning strategies to infer MBTI personality types from written texts. These approaches range from traditional feature extraction methods like TF-IDF and bag-of-words to more sophisticated machine learning and deep learning frameworks. Scholars have investigated various language attributes, such as term frequency patterns, grammatical structures, and semantic vector representations, to uncover subtle indicators of personality traits embedded in text-based communication.

Sheth and Pandhare [4], as well as Ontoum and Chan [5], explored the use of diverse algorithms—including Naive Bayes, Support Vector Machines, Random Forest, Decision

Trees, and k-Nearest Neighbours—to analyse social media content for personality detection, highlighting that the performance of these approaches can vary depending on the model applied. Their findings revealed promising applications in recruitment and personalized digital interactions. Furthermore, the research by Vaddem and Agarwal [6] showcased the strong capability of XGBoost in pinpointing significant linguistic features, delivering impressive accuracy rates, especially when applied to texts in languages other than English. Their findings emphasize how advanced models such as XGBoost and SVM can effectively adapt to perform nuanced personality analysis across diverse linguistic contexts. Moreover, Amirhosseini and Kazemian [7] developed a novel methodology leveraging meta programs and extreme gradient boosting (XGBoost). They compared the performance of their model with to other existing methods and the results demonstrated enhanced accuracy and reliability in personality prediction, with real-life applications for psychologists and neuro-linguistic programming (NLP) practitioners.

Building on these foundational insights, pre-trained language models such as BERT have transformed personality prediction accuracy. Kaushal et al. [8] employed BERT for sentiment analysis and MBTI prediction, emphasizing its ability to enhance contextual understanding. Among tested classifiers, SVM emerged as the most accurate. Christian et al. [9] advanced this further by integrating BERT, RoBERTa, and XLNet in a multi-model architecture, achieving an impressive 88.5% accuracy and an F1 score of 0.882 by utilizing data from diverse social media platforms.

Other studies have highlighted innovative applications of BERT-based models. Guo et al. [10] analysed open-domain and task-oriented dialogues, finding personality predictions to be more accurate in open-domain contexts. In a comparable vein, Garcia Dos Santos and Paraboni [11] illustrated that carefully optimised BERT architectures surpassed conventional techniques such as bag-of-words models and fixed word embeddings, with their superiority confirmed through comprehensive cross-validation experiments.

Further research has explored ensemble and boosting techniques to enhance model performance. Shafi et al. [12] compared multiple classifiers, identifying Ensemble Bagged Trees as a standout with training accuracy of 98.4% and test accuracy of 70.75%. Mushtaq et al. [13] utilized gradient boosting and K-means clustering to predict personality types, emphasizing the authenticity of social media data over traditional survey methods. Their approach significantly outperformed naive Bayes classification. Recent advancements also address challenges like data imbalance. Ryan et al. [14] integrated logistic regression, SVMs, boosting methods, Word2Vec embeddings, and SMOTE to improve F1 scores, showcasing the impact of balanced datasets on predictive accuracy. These studies collectively highlight the evolving landscape of machine learning applications in personality prediction, underscoring the potential of advanced algorithms and pre-trained language models to deliver meaningful psychological insights.

B. Research Gap

The research highlights gaps in machine learning and deep learning approaches to personality prediction, emphasizing the need for head-to-head comparisons of models on identical datasets to identify their strengths and limitations for applications like psychology, career counseling, and marketing. Previous studies have heavily focused on the Big Five personality framework, neglecting the MBTI due to limited datasets. Additionally, basic optimization techniques like GridSearch and lack of attention to label imbalance and inconsistent text sample lengths have hindered model performance.

This research aims to address these gaps by utilizing a new large and diverse dataset for MBTI personality types and conducting rigorous comparative analyses between the best performing machine learning and deep learning models which were previously introduced in other studies, as well as new models which were not tried before for personality prediction task based on MBTI. All models will be optimised based on a Bayesian optimisation framework for hyperparameter tuning and the label imbalance issue will be addressed in the new dataset. All the samples will also be reasonably restricted to ensure that there will be sufficient number of characters in the text for model training. This effort will not only enhance the predictive accuracy and generalizability of personality prediction models but also ensure their applicability in a wide range of real-world scenarios.

III. Methodology

A. Dataset

The dataset employed in this study comprises a wide-ranging assortment of social media content, including posts from online forums, Twitter feeds, and blog articles. Each entry in the dataset is linked to the respective author's MBTI personality classification. This extensive dataset includes more than 110,000 individual records, providing a robust foundation for analysis. Table 1 presents examples of these social media entries alongside the corresponding MBTI type assigned to each author.

TABLE I. Sample of social media posts and the MBTI personality type labels

Social Media Post	MBTI Label
"I fully believe in the power of being a protector, to give a voice to the voiceless. So, in that spirit I present this film, and hope it received in the spirit of compassion ..."	INFJ
"He doesn't want to go on the trip without me, so me staying behind wouldn't be an option for him. I think he really does believe that I'm the one being unreasonable. He continues to say that..."	ENFP
"Fair enough, if that's how you want to look at it. Like I stated before, they were incredibly naive in their comments. However, they think those are things that would help us because those are the..."	INTJ
"I love feeling affectionate for the one I love and care for. I	ISFJ

care about her very deeply. I want to protect her, make her happy, and always be by her side. I feel that romantic love is probably...”	
“I have a cousin with Aspergers. He is really a great kid. Right now, he is going to engineering school. When we were kids we never really talked, it was mostly just smiling and playing games…”	ISTP

The dataset presents numerous complexities for personality classification, such as inconsistent and disorganised text, diverse writing styles, and the nuanced presence of elements like sarcasm, irony, and ambiguous expressions. To address these challenges and derive valuable features from the raw textual content, a structured preprocessing pipeline was developed, as detailed in the subsequent sections.

B. *Exploratory Data Analysis (EDA)*

Performing Exploratory Data Analysis (EDA) is a fundamental step in this research, as it provides essential insights into the underlying structure of the dataset and clarifies the distribution trends of various personality types. Figure 1 depicts the spread of MBTI categories in the dataset employed for this study. The visualisation highlights notable discrepancies in the prevalence of different personality types. Specifically, the INTP type emerges as the most dominant, exceeding 24,000 records, while types like ESFP and ESFJ appear far less frequently, indicating their minority status within the data. This significant imbalance draws attention to a key challenge inherent in the dataset — the uneven distribution of personality classes. With certain types being heavily represented and others scarcely present, the imbalance could potentially hinder the generalisability and fairness of predictive models. To build reliable and unbiased classifiers, it becomes essential to implement strategies that address and mitigate this class distribution issue.

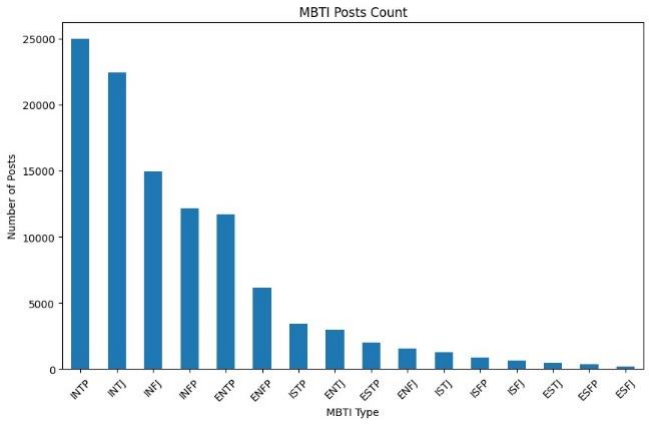

Figure 1: Distribution of personality types.

C. *Data Preprocessing*

This process involved defining English stop words and initializing a lemmatizer using the Natural Language Toolkit (NLTK) Python library. These preprocessing steps aimed to streamline the textual data by removing non-informative words and normalizing word forms to their base representation. Moreover, a custom preprocessing function, *preprocess_text*, was implemented to clean and standardize the textual data. This function tokenized the text, removed non-alphabetic tokens, and excluded predefined stop words, followed by lemmatization to ensure linguistic consistency. The processed text was then converted into a structured format suitable for machine learning analysis through the application of the Term Frequency-Inverse Document Frequency (TF-IDF) vectorizer. To enhance computational efficiency and focus on the most informative features, the vectorizer was set to limit the feature space to the top 5,000 terms based on their relevance.

Additionally, categorical target variables were transformed using one-hot encoding technique to convert them into a binary matrix representation. This transformation is essential for machine learning and deep learning algorithms, that require numerical input and cannot inherently process categorical data. These preprocessing steps ensured that the dataset is appropriately formatted and optimized for subsequent machine learning and deep learning model development and evaluation.

D. *Addressing Class Imbalance Issue*

To address the imbalance in class representation revealed during the exploratory analysis, this study employed the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE enhances the dataset by synthesising new instances for the minority classes, effectively equalising the distribution across all personality types. By incorporating this method, the model's capacity to generalise to less frequent personality categories is expected to improve significantly. Following the application of SMOTE, the dataset is transformed into a more balanced form, making it suitable for the subsequent machine learning training processes. This step ensures that the predictive models do not exhibit bias toward the more dominant classes and are instead capable of delivering reliable performance across the full range of MBTI personality classifications.

E. *Splitting the Dataset*

In this study, the dataset was partitioned into training and testing subsets using the train_test_split method provided by the scikit-learn library in Python. Specifically, 70% of the data was allocated for training the models, while the remaining 30% was reserved to evaluate their performance.

F. *Model development*

As outlined in Section II, a key knowledge gap identified in the literature is the lack of comprehensive analysis comparing the performance of machine learning and deep learning models for personality prediction on an identical dataset. To address this gap, this research aimed to implement some of the best-performing models identified in previous studies—originally tested on diverse datasets—alongside the development of novel models that have not yet been explored for this task.

Eight machine learning and deep learning models were developed in this work including (i) Support Vector Classifier (SVC), (ii) Logistic regression (LR), (iii) Light Gradient Boosting Machine (LightGMB), (iv) Naïve Bayes, (v)

Random Forest (RF), (vi) Extreme Gradient Boosting (XGBoost), (vii) Multi-Layer Perceptron Neural Network (MLP), and (viii) Bidirectional Encoder Representations from Transformers (BERT). A thorough comparative analysis was then conducted to evaluate and benchmark these models, providing a more robust understanding of their relative strengths and weaknesses in personality prediction.

(i) Support Vector Classifier (SVC):

This model is particularly well-suited to the high-dimensional nature of text data. This model was selected due to its strong performance in text classification tasks and its demonstrated robustness against overfitting, even when dealing with the large feature sets typically associated with textual data. SVC constructs a hyperplane in a high-dimensional space to classify data points. It uses the kernel trick to enable non-linear separations while maximizing the margin between classes. The decision boundary is defined as:

$$f(x) = w^T x + b$$

Where w is the weight vector, x is the input vector, and b is the bias term. SVC minimizes the hinge loss based on this formula:

$$L(w,b) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} \max(0, 1 - y_i(w^T x_i + b))$$

This model map data to a high-dimensional feature space using kernels and then solves the optimization problem to find the hyperplane that maximizes the margin. It will then classify points based on their distance to the hyperplane. SVC's ability to handle sparse data, such as that represented by TF-IDF vectors, makes it an ideal choice for this analysis.

(ii) Logistic regression (LR):

This model is well-suited for the text-based personality prediction task, as it estimates the probabilities associated with different personality types. Logistic regression models the probability of a class label using a logistic function. It is particularly effective for linearly separable data. The logistic function is given by:

$$P(y = 1|x) = \frac{1}{1 + e^{-(w^T x + b)}}$$

The loss function is the binary cross-entropy:

$$L(w,b) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)]$$

This model, in the first step, initializes weights and bias. Then it computes the probability of each class using the logistic function. Finally, it optimizes the weights and bias using gradient descent to minimize cross-entropy loss. The algorithm's use of a logistic function enables it to handle non-linear relationships, a critical advantage given the inherent complexities of language data.

The interpretability and simplicity of Logistic Regression are particularly advantageous, making it an effective baseline for comparison with more sophisticated models while providing insights into the relationships between linguistic features and personality traits.

(iii) Light Gradient Boosting Machine (LightGMB):

The LightGBM model is particularly well-suited for text-based personality prediction tasks due to its efficiency and ability to handle large datasets with numerous features. This model uses a gradient-boosting framework and builds trees leaf-wise rather than level-wise, reducing loss more efficiently. LightGBM optimizes the following objective:

$$L = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Where ℓ is the loss function and Ω is the regularization term. The model will be initialized with a constant prediction. Then it computes residuals for the current model. The model then fits a weak learner (tree) to the residuals and updates predictions and repeats until convergence.

LightGBM's ability to manage high-dimensional data and its inherent regularization mechanisms make it robust against overfitting, which is especially valuable in text analysis. Additionally, the model's flexibility and scalability provide a strong foundation for achieving high predictive performance, making it an excellent choice for this application.

(iv) Naïve Bayes:

The Naïve Bayes classifier was also explored due to its well-established effectiveness in text classification tasks, particularly for its simplicity, computational efficiency, and scalability when working with large datasets. This model assumes conditional independence between features. It uses Bayes' theorem to compute probabilities:

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

Naïve Bayes assumes:

$$P(X|y) = \prod_{i=1}^{n} P(x_i|y)$$

The model computes prior probabilities for each class and then Computes likelihoods of features given a class. It then apply Bayes' theorem to compute posterior probabilities and assigns the class with the highest posterior probability. Its probabilistic approach makes it especially suitable for text data, where features such as term frequencies play a critical role.

(v) Random Forest (RF):

This model is highly effective for the text-based personality prediction task, as it utilizes an ensemble learning approach to create a collection of decision trees, each contributing to the final prediction. Random Forest minimizes:

$$E = \frac{1}{T}\sum_{t=1}^{T} E_t$$

Where T is the number of trees and E_t is the error of the t-th tree. Random Forest's ability to handle non-linear relationships and capture intricate patterns within the data makes it particularly suited for complex text datasets. The algorithm's robustness to overfitting, achieved through techniques like bootstrapping and feature randomness, ensures reliable performance even with noisy or unbalanced data.

Additionally, Random Forest provides valuable insights into feature importance, highlighting the linguistic attributes most strongly associated with personality traits. This interpretability, combined with its high accuracy, establishes Random Forest as a powerful baseline for comparison with more advanced models.

(vi) Extreme Gradient Boosting (XGBoost):

XGBoost enhances the traditional gradient boosting framework by incorporating advanced regularization techniques (L1 and L2), efficient handling of sparse data, and parallel processing capabilities. These improvements make XGBoost highly effective for predictive modelling tasks, particularly in high-dimensional datasets such as text-based personality prediction. This model minimizes the following objective function:

$$\mathcal{L} = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Where $\ell(y_i, \hat{y}_i)$ is the loss function that measures the difference between the predicted value $\hat{y}_i$ and the actual value y_i. Commonly used loss functions include mean squared error (MSE) for regression and log-loss for classification. $\Omega(f_k)$ is the regularization term that penalizes the complexity of the model to avoid overfitting. The regularization term is given by:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

Where T is the number of leaves in the tree, w_j is the weight of leaf j, γ is the minimum loss reduction required to make a further partition on a leaf node, and λ is the regularization parameter that controls the L2 penalty on leaf weights.

This model **s**tarts with a constant prediction, often the mean value of the target variable. Then in each iteration, add a tree to minimize the residual error. The model then uses the gradient (first derivative) and Hessian (second derivative) of the loss function to guide tree construction. It finally adjusts predictions by combining previous predictions with the current tree's output.

XGBoost excels at handling non-linear relationships and complex interactions within the data, making it particularly well-suited for nuanced text datasets. Its ability to manage sparse data and incorporate regularization techniques minimizes overfitting, ensuring robust generalization across diverse samples. Furthermore, XGBoost provides interpretable insights into feature importance, allowing identification of key linguistic patterns associated with personality traits. These attributes, coupled with its efficiency and scalability, make XGBoost a strong contender for comparison with other advanced models in personality prediction.

(vii) Multi-Layer Perceptron Neural Network (MLP):

Beyond conventional machine learning techniques, this research incorporated a deep learning strategy by deploying a multi-layer perceptron (MLP) neural network to predict personality types from textual inputs. The MLP architecture comprises densely connected layers, enabling the capture of complex, non-linear patterns within the data. The hidden layers employ the Rectified Linear Unit (ReLU) activation function to enhance learning efficiency and model expressiveness:

$$f(x) = \max(0, x)$$

The architecture of implemented model in this study consists of an input layer followed by three hidden layers with 512, 256, and 128 neurons, respectively. Dropout regularization is applied after each hidden layer to mitigate overfitting. The network concludes with an output layer utilizing a SoftMax activation function, which produces a probability distribution across the personality type classes. The SoftMax activation function uses this formula:

$$P(y = i|x) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_i}}$$

In this formula, $P(y = i|x)$ is the probability of the class i being the correct class, given the input x. This is the predicted probability for class i. z_i is the logit or raw score for class i, output by the network's final layer before applying the SoftMax function. e^{z_i} converts the raw score z_i into a positive number. This step ensures that probabilities are always non-negative. $\sum_{j=1}^{n} e^{z_i}$ is the sum of the exponentiated scores for all n classes. This serves as a normalizing factor to ensure that the probabilities across all classes sum to 1. Finally, dividing e^{z_i} by the sum ensures that the resulting value is a valid probability (between 0 and 1).

The model was compiled using the categorical cross-entropy loss function, ideal for multi-class classification problems, and the Adam optimizer, chosen for its adaptive learning rate and computational efficiency. The training process was conducted over 50 epochs with a batch size of 64, and performance metrics were monitored on both training and validation datasets to ensure effective learning and generalization. The combination of this carefully designed architecture, dropout regularization, and the SoftMax output layer allowed the model to handle class imbalance and prevent overfitting effectively, resulting in accurate predictions of personality types from text data.

(viii) Bidirectional Encoder Representations from Transformers (BERT):

The Bidirectional Encoder Representations from Transformers (BERT) model is exceptionally well-suited for the text-based personality prediction task due to its deep

contextual understanding of language. By processing input text bidirectionally, BERT captures the nuanced relationships between words and their surrounding context, which is critical for analysing the complexities of linguistic patterns associated with personality traits. Its pre-trained transformer architecture allows it to leverage vast amounts of textual data, enabling fine-tuning on the task-specific dataset for enhanced performance.

BERT processes text bidirectionally, meaning it considers both the preceding and succeeding context of a word. This enables it to capture deep semantic and syntactic relationships, making it effective for text-based personality prediction. Each input token is represented as a combination of embeddings:

$$Input_i = E_{token_i} + E_{segment_i} + E_{position_i}$$

Where E_{token_i} is word embedding for token i, $E_{segment_i}$ is embedding indicating sentence membership ([SEP] separates segments), and $E_{position_i}$ is positional embedding indicating the token's position in the sequence. For each token, BERT calculates attention scores to determine its importance relative to other tokens:

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

Q, K, and V are representing Query, Key, and Value matrices derived from token embeddings. d_k is the dimensionality of the key vectors (used for scaling). BERT applies multiple transformer layers, where each layer consists of:

Multi-Head Attention:

$$MultiHead(Q,K,V) = Contact(head_1, \dots, head_h)W^Q$$

Feed-Forward Network:

$$FFN(x) = ReLu(xW_1 + b_1)W_2 + b_2$$

Applies a two-layer neural network to each token embedding. BERT outputs a contextual embedding for the [CLS] token, summarizing the entire sequence. The [CLS] embedding is passed to a classifier for personality prediction:

$$P(y|x) = softmax(W.h_{[CLS]} + b)$$

Where $h_{[CLS]}$ is the final embedding of the [CLS] token. W and b are trainable weights and bias for the classification layer. This model tokenizes text and adds special tokens ([CLS], [SEP]), and then passes tokens through transformer layers to compute contextual embeddings. It will then extract the [CLS] embedding and fine-tune the model using labelled data with a classification head. This approach allows BERT to leverage its bidirectional contextual understanding for nuanced personality prediction.

The model's ability to learn rich semantic representations and handle intricate dependencies within text makes it a powerful tool for personality prediction, surpassing traditional methods in capturing subtle linguistic cues that define individual differences. BERT's state-of-the-art capabilities provide a robust foundation for comparison with other machine learning and deep learning models while offering deep insights into the interplay between language use and personality.

G. Model Optimization

A Bayesian optimization framework was employed in this study using *Optuna* Python library for hyperparameter tuning and improving the performance of the implemented models. Unlike conventional optimization methods such as GridSearch—identified as a limitation in previous literature—Bayesian optimization framework offers a more efficient and robust strategy for both traditional machine learning and deep learning models.

This approach intelligently balances exploration and exploitation, enabling faster convergence to optimal parameter values while reducing computational overhead. Table 2 presents the optimal values identified for each parameter across the different models.

TABLE II. HYPERPARAMETER TUNING OUTCOMES

Model	Parameter	Optimum Value
SVC	*C*	*1.8336*
	kernel	*rbf*
	gamma	*0.0491*
	shrinking	*True*
	tol	*0.001*
	cache_size	*200*
	random_state	*42*
LR	C	0.01
	max_tier	1000
	solver	'saga'
LightGBM	boosting_type	gbdt
	objective	multiclass
	metric	multi_logloss
	num_leaves	30
	max_depth	9
	min_data_in_leaf	20
	learning_rate	0.1
	num_iterations	100
Naïve Bayes	var_smoothing	1e-9
RF	n_estimators	130
	max_depth	19
	min_samples_split	7
	min_samples_leaf	3
	bootstrap	True
XGBoost	n_estimators	423
	max_depth	9
	learning_rate	0.1
	subsample	0.5
	colsample_bytree	0.9
MLP	Hidden Layer Size	**Layer 1:** 512 neurons **Layer 2:** 256 neurons **Layer 3:** 128 neurons
	Activation Function in Hidden Layers	relu
	Activation Function in Output Layer	SoftMax

	solver	adam
	Learning Rate	0.001
	Max Iterations	50
	Batch Size	64
BERT	hidden_size	768
	num_hidden_layers	12
	num_attention_heads	12
	intermediate_size	3072
	hidden_act	gelu
	dropout_prob	0.1
	learning_rate	1e-5
	weight_decay	0.01
	adam_epsilon	1e-8
	batch_size	16
	epochs	4
	max_seq_length	256
	do_lower_case	True
	strip_accents	None

IV. Results

The evaluation metrics—accuracy, F1 score, recall, and precision—were computed for eight implemented models to assess their performance in predicting personality types from text data. The results are summarized in Table 3.

TABLE III. Prediction results for each model

Model	Accuracy	F1 Score	Recall	Precision
SVC	0.85	0.86	0.85	0.86
LR	0.82	0.82	0.82	0.83
LightGBM	0.68	0.69	0.68	0.70
Naïve Bayes	0.69	0.69	0.69	0.70
RF	0.84	0.84	0.84	0.85
XGBoost	0.86	0.86	0.86	0.86
MLP	0.80	0.80	0.80	0.80
BERT	0.93	0.93	0.93	0.93

According to Table 3, BERT model emerged as the top-performing model, achieving the highest accuracy of 93%, along with a consistent F1 score, recall, and precision of 0.93 across all metrics. This indicates its superior ability to handle textual data and accurately predict personality types.

XGBoost and SVC demonstrated comparable performances, with XGBoost achieving an accuracy of 86% and SVC following closely at 85%. Both models achieved an F1 score, recall, and precision above 0.85, reflecting their strength in capturing complex patterns within the dataset. Random Forest (RF) performed slightly lower than XGBoost and SVC, with an accuracy of 84% and F1 score, recall, and precision values of 0.84 and 0.85, respectively. It remains a robust alternative but falls short of the gradient boosting approaches. Logistic Regression (LR) and Multilayer Perceptron (MLP) delivered moderate performances, with accuracies of 82% and 80%, respectively. While both models provided consistent results across all metrics, they were outperformed by ensemble and advanced machine learning models. Naïve Bayes and LightGBM displayed the lowest performances, achieving accuracies of 69% and 68%, respectively. Their F1 scores, recall, and precision were also lower compared to the other models, indicating limitations in their ability to capture the intricacies of textual data.

These results highlight the superiority of deep learning models and advanced ensemble methods for text-based classification tasks, while traditional models provided solid baseline comparisons.

V. Discussion

The results demonstrate a clear hierarchy in model performance, with BERT outperforming all other models. This is attributed to its sophisticated transformer architecture, which excels at understanding contextual and semantic nuances in textual data. The consistent metrics across accuracy, F1 score, recall, and precision further underline BERT's robustness in handling text classification tasks, particularly in high-dimensional spaces. The BERT model developed in this research also significantly outperformed previous works, such as that of Christian et al. [9], who integrated BERT, RoBERTa, and XLNet into a multi-model architecture. Their approach achieved an accuracy of 88.5% and an F1 score of 0.88 using data from social media platforms.

XGBoost and SVC also performed remarkably well, achieving high accuracy and precision levels. These models effectively balance complexity and efficiency, making them suitable for tasks requiring high performance without the computational intensity of deep learning models. XGBoost, with its gradient boosting framework, demonstrated its ability to capture intricate patterns in the data, while SVC's margin maximization capabilities proved effective for classification.

Random Forest offered solid results, though slightly behind XGBoost and SVC. While it provides interpretability and robustness, its relatively lower performance highlights the limitations of random forest models compared to gradient boosting techniques in handling complex relationships in high-dimensional datasets. Logistic Regression (LR) and Multilayer Perceptron (MLP) served as effective baselines but were outperformed by ensemble and deep learning models. The moderate performance of MLP suggests that further optimization and a larger dataset may be necessary for neural networks to fully achieve their potential in this context. Naïve Bayes and LightGBM struggled with the dataset, achieving the lowest scores among all models. Naïve Bayes, constrained by its assumption of feature independence, likely failed to capture the intricate relationships in the textual data. Similarly, LightGBM, while powerful for structured datasets, appears less suited for unstructured text data in its current configuration.

These findings emphasize the critical role of model selection based on dataset characteristics. While deep learning models like BERT provide unmatched performance, their computational requirements may limit their applicability in resource-constrained environments. Ensemble methods like XGBoost and SVC offer a practical alternative with strong results and relatively lower resource demands. Finally, addressing the observed class imbalance remains essential for improving generalizability across personality types. The use of SMOTE in this study ensured balanced training data, which likely contributed to the improved performance of models like SVC and XGBoost.

Future research should explore the integration of multilingual datasets to enhance the generalizability of these models across various populations and provide further

insights into the models' strengths and limitations, guiding their optimal application in real-world personality prediction. Incorporating additional features, such as demographic information or social network interactions, could also be explored in future works to further enhance the accuracy and robustness of personality prediction models. Demographic attributes, including age, gender, and cultural background, can provide contextual insights that complement linguistic patterns, enriching the understanding of personality traits. Similarly, data derived from social network interactions—such as likes, shares, and comments—may capture behavioural nuances and interpersonal dynamics that are not readily apparent from textual data alone. By integrating these multidimensional features into machine learning and deep learning frameworks, future studies can develop more holistic and context-aware personality prediction models, thereby improving their applicability across diverse real-world scenarios.

Finally, as the BERT model in this study has demonstrated superior performance compared to other models, researchers with limited computational resources can adapt or optimize BERT for personality prediction by employing strategies such as knowledge distillation, model pruning, and quantization. These techniques reduce the model's size and computational requirements while maintaining its predictive accuracy. Leveraging smaller pre-trained models, efficient training techniques, and deployment optimizations further enhances the usability of BERT in resource-constrained settings. Such approaches democratize advanced NLP capabilities, enabling a broader range of researchers and practitioners to access and utilize these powerful tools effectively.

VI. Conclusion

This research makes a significant contribution to the field of personality prediction by addressing several key limitations in prior studies and advancing the methodological rigor in this domain. First, it fills a critical gap by conducting a direct comparative analysis of machine learning and deep learning models on identical datasets, providing actionable insights into the relative strengths and weaknesses of each algorithm. This approach enables researchers and practitioners to make more informed decisions about model selection and optimization for specific applications.

Second, by focusing on the MBTI framework—a previously underutilized personality model in machine learning research due to data constraints—this study broadens the scope of personality assessment beyond the commonly used Big Five framework. The use of a large-scale, diverse dataset with balanced labels and sufficient text length ensures that the findings are robust, generalizable, and reflective of real-world conditions.

Third, this research leverages Bayesian optimization for hyperparameter tuning, overcoming the inefficiencies of traditional methods like GridSearch, particularly for complex deep learning models. By addressing the label imbalance issue and standardizing textual input length, the study ensures fair and effective training for all models.

The results highlight the superior performance of deep learning models, with BERT achieving the highest accuracy (93%), followed by XGBoost (86%), SVC (85%), and RF (84%). The BERT model also significantly outperformed the models implemented in previous works in this field. These findings underscore the importance of advanced models for handling nuanced and unstructured textual data. In contrast, simpler models like Naïve Bayes and LightGBM were less effective in this context, reaffirming the need for tailored approaches to different datasets and applications.

By addressing these methodological gaps, this study not only enhances the predictive accuracy and generalizability of personality prediction models but also lays the groundwork for future research. The insights gained from this work can inform the development of more robust, scalable, and versatile personality prediction systems applicable across diverse fields, including psychology, education, and marketing.

References

[1] Myers, S. (2016) Myers-Briggs typology and Jungian individuation, Journal of Analytical Psychology, 61(3), 289–308. Doi: 10.1111/1468-5922.12233.

[2] Stein, R. and Swan, A. (2019). Evaluating the validity of Myers-Briggs Type Indicator theory: A teaching tool and window into intuitive psychology, Soc Personal Psychol Compass. Dio: 10.1111/spc3.12434.

[3] Pittenger, D.J. (1993). The Utility of the Myers-Briggs Type Indicator, Review of Educational Research, 63(4), 467-488. Dio: 10.2307/1170497.

[4] Sheth, A. V., & Pandhare, D. R. (2024). *Myers-briggs Personality Prediction using Machine Learning Techniques.* International Journal for Multidisciplinary Research.

[5] Ontoum, S., & Chan, J. (2022). Personality Type Based on Myers-Briggs Type Indicator with Text Posting Style by using Traditional and Deep Learning. Bangkok: University of Technology Thonburi.

[6] Vaddem, N., & Agarwal, P. (2020). Myers Briggs Personality Prediction using Machine.

[7] Amirhosseini, M. H., & Kazemian, H. (2020). *Machine Learning Approach to Personality Type Prediction Based on the Myers–Briggs Type Indicator.* London: School of Computing and Digital Media, London Metropolitan University.

[8] Kaushal, P., Bharadwaj, N., Koundinya, A., Pranav, & Koushik. (2021). *Myers-briggs Personality Prediction and Sentiment Analysis of Twitter using Machine learning Classifiers and BERT.* International Journal of Information Technology and Computer Science (IJITCS).

[9] Christian, H., Suhartono, D., Chowanda, A., & Zamli, K. (2021). Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. Journal of Big Data.

[10] Guo, A., Hirai, R., Ohashi, A., Chiba, Y., Tsunomori, Y., & Higashinaka, R. (2024). *Personality prediction from task-oriented and open-domain human–machine dialogues.* natureportfolio.

[11] Dos Santos, V. G., & Paraboni, I. (2022). Myers-Briggs personality classification from social media text using pre-trained language models.

[12] Shafi, H., Sikender, A., Jamal, I. M., Ahmad, J., & Aboamer, M. A. (2021). *A Machine Learning Approach for Personality.* Journal of Independent Studies and Research Computing Vol 19 Issue 2.

[13] Mushtaq, Z., Ashraf, S., & Sabahat, N. (2020). *Predicting MBTI Personality type with K-means.* IEEE.

[14] Ryan, G., Katarina, P., & Suhartono, D. (2023). MBTI Personality Prediction Using Machine Learning and SMOTE for Balancing Data Based on Statement Sentences. Information.